



# PERSONALITY AND INTELLIGENCE

A STUDY OF SOME RESPONSES OTHER THAN  
INTELLECTUAL NOTED IN A SIMPLE  
MENTAL TEST SITUATION

BY

JANET FOWLER NELSON

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE FACULTY  
OF PHILOSOPHY, COLUMBIA UNIVERSITY

BUREAU OF PUBLICATIONS  
*Teachers College, Columbia University*  
NEW YORK CITY  
1931

Copyright, 1931, by TEACHERS COLLEGE, COLUMBIA UNIVERSITY

TO MY FATHER  
EDGAR CRAYTON FOWLER  
THIS BOOK IS GRATEFULLY  
AND LOVINGLY DEDICATED

Digitized by the Internet Archive  
in 2017 with funding from

This project is made possible by a grant from the Institute of Museum and Library Services as administered by the Pennsylvania Department of Education through the Office of Commonwealth Libraries

## ACKNOWLEDGMENTS

The preliminary study of techniques on which this research is based was made possible through the coöperation and interest of Dr. Helen Thompson Woolley, formerly Director of the Child Development Institute, Teachers College, Columbia University, Mrs. Kathleen Ewell, Director of the Waverly School, New York City, and members of the staff of the Child Development Institute, to all of whom the writer is grateful.

The mailing list used as a basis for the selection of subjects was obtained through the courtesy of the Board of Health of White Plains, New York.

Most gratefully does the writer acknowledge the supervision and criticism given by Professors Rudolph Pintner and Ralph B. Spence, of Teachers College, who have been consulted throughout the course of the study and whose critical comments have been invaluable.

The writer is especially indebted to Dr. Dorothy Swaine Thomas not only for the comments and suggestions which she made throughout the progress of the present study, but also because it was during an earlier association with Dr. Thomas that the point of view expressed in this research developed into a tentative methodology which has been subsequently refined and employed.

To Clarence E. Nelson, the writer wishes to express most sincere appreciation for his assistance in the preparation of the data for publication and for his personal encouragement and interest which have greatly facilitated the progress of the whole study.

J. F. N.



## CONTENTS

CHAPTER	PAGE
I. INTRODUCTION .....	I
II. DEVELOPMENT OF METHOD .....	5
III. PLAN OF MAIN STUDY .....	12
IV. RESISTANCE .....	19
V. PERSISTENCE .....	31
VI. SPONTANEOUS CONVERSATION .....	36
VII. SPONTANEOUS ACTIVITY (GROSS) .....	41
VIII. "INITIATIVE" INTEREST IN TEST MATERIALS .....	44
IX. SPONTANEOUS USE OF TEST MATERIALS .....	47
X. SEX DIFFERENCES .....	50
XI. INTERRELATIONSHIPS EXISTING AMONG THE VARIOUS MEASURES .....	52
XII. INDIVIDUAL DIFFERENCES .....	55
XIII. SUMMARY .....	60
BIBLIOGRAPHY .....	62



## CHAPTER I

### INTRODUCTION

Techniques for the psychological testing of children at the preschool level are being rapidly developed at the present time. They still cannot compare in stability and in standardization, however, with the better established techniques for testing children at higher chronological age levels. The traditional teacher-pupil attitudes which usually carry over and facilitate the testing situation at school-age levels are found to be non-existent when one is dealing with children of preschool age. We frequently find in such children dependence on the examiner for motivation and encouragement and, at the same time, rather startling evidence of independence of thought and activity. There is on the part of the child no conception of being tested, but instead the "game" aspect of the situation prevails, with all its implications of spontaneity and interest. The very fact that comparatively few small children will remain quietly seated at the testing table throughout the time required for a psychological test introduces a problem of adjustment to testing techniques and standardization of procedure which has to be recognized and met. In addition, we are confronted with the child's enthusiasms and chance interests, and with his lack of inhibition in his emotional reactions to the total situation.

In an earlier study the writer discussed these aspects of the problem of standardization and reached the conclusion that "up to a certain point it is undoubtedly to be welcomed—standardization of material and of procedure, objective scoring, etc. However, until techniques are developed for evaluating personality differences, thereby serving as a control, too rigid

control of the continuity of the test situation, for instance, would seem to imply false standards of standardization. Testing, especially at the lower chronological age levels, involves two large variables—personality of the child and of the tester. Failure to recognize and provide for these variables is not the premise for true standardization."<sup>1</sup>

The standardization of test situations with human beings as subjects involves many difficulties. Even if the greatest care is taken in setting up the situation, it is undoubtedly fallacious to assume that "other conditions" are equalized. Rigidity of procedure may actually serve further to confuse intelligence, as such, with personality reactions. It has therefore seemed important to attack the problem from a somewhat different standpoint from any which is usually taken.

The extent of the personality variation of different subjects in situations supposedly standardized to measure intelligence has been studied by recording and analyzing certain simple and clearly defined personality reactions without in any way interfering with the major purpose of the mental test. Thus resistance to the examiner and to the test situation has been studied, that is, frequency of occurrence of resistance in defined units, individual differences as to resistance among children, the relationship of resistance to specific test items, etc. General problems, such as a comparison of the average mental age of more resistant children with that of less resistant children, have been approached. The same procedure has been followed in studying the spontaneous activity of the child. It has been the purpose of this study to find out, first, whether activity as such is correlated with intelligence test ratings as they are now obtained, and, second, whether the child is revealing by his activity a type of behavior reaction in a more or less controlled situation, which perhaps would not be reflected in the intelligence test score, but which may have importance for the study of personality differences. And if so, can this behavior

<sup>1</sup> Thomas, Dorothy Swaine, and Associates, *Some New Techniques for Studying Social Behavior*. Chapter on "Preliminary Report on Some Uses of the Psychological Test Situation for Studying Personality Differences," p. 119.

reaction be measured? Therefore, throughout this study, the total psychological test situation has been considered from the point of view of children's responses other than those called for by the test itself. A technique for the objective measurement of these reactions has been developed. Data are presented on responses indicating resistance, spontaneous activity (gross), spontaneous conversation, initiative, and persistence, as revealed in the total psychological test situation. The study indicates group norms and emphasizes the interrelationships of these behavior patterns.

It is unfortunate that much incidental behavior, interesting and psychologically significant, has had to be omitted from consideration because the techniques developed for recording it were considered unreliable. The selection of the test items to be studied and the reasons for the omission of others will be discussed in Chapter II.

To obtain a really adequate picture of personality responses to a psychological test situation, the variations between examiners as well as those between subjects should be studied. To assume that the one examiner in this study showed no variation in her reactions from one situation to another would be over-optimistic. The development of techniques for approaching these factors has, however, been projected into a future study; they could not be included in this one.

Besides that furnished by the current interest in techniques for testing at the preschool level, another impetus for the present study is provided by the fact that, regardless of the effect of incidental behavior on intelligence quotients as such, interest is being increasingly directed to an understanding of the "whole child." Responding to this interest, one finds in the literature, especially at higher age levels, an emphasis on the development of tests and measures of personality traits. Before complicating testing programs for children at the preschool level with batteries of many tests, it seems logical to determine whether or not we have the opportunity of studying some personality traits as they are revealed in a simple intelli-

gence test situation. In dealing with children of preschool age the opportunity to do this seems admirable. It is the writer's impression and opinion that little children reveal much that is interesting and significant even in a so-called controlled situation. Certainly, anyone who has given intelligence tests to children of nursery school age comes from the psychological laboratory with more or less clearly defined impressions of a given child's personality. The writer hopes to present a few methods of banishing such impressions from the field of mere conjecture and to provide a means of evaluating them more objectively.

## CHAPTER II

### DEVELOPMENT OF METHOD

The purpose of this study, as noted in the Introduction, is to devise a means of evaluating the spontaneous activity and conversation, incidental to the psychological test situation, of a group of three-year-old children.

A preliminary study was first made of seventeen nursery school children. For this privilege the writer is indebted to Dr. Helen Thompson Woolley, formerly Director of the Child Development Institute of Teachers College, Columbia University, and to Mrs. Kathleen Ewell, Director of the Waverly School, New York City. The psychological tests given were the Kuhlmann Revision of the Binet scale<sup>1</sup> and the Merrill-Palmer Performance Scale<sup>2</sup>—the tests which were subsequently used in the main study.

The original plan was for the tester to check on the test blank the test items resisted by the children. Resistance was arbitrarily defined, as in a previous study,<sup>3</sup> as (1) the ignoring of material presented or question asked; (2) verbal resistance such as "No," "I won't"; (3) physical resistance, manifesting itself in walking or running from the table or room, or in definitely pushing away proffered material; (4) certain instances of "I can't" or "I don't know." This last criterion is equivocal, for undoubtedly a large majority of such responses indicate a child's insight into his own limitations. For instance, "I don't know" appearing in response to one of a list of questions, such as those in the Action Agent test in the Merrill-Palmer Scale, preceded and followed by actual responses,

<sup>1</sup> Kuhlmann, Fred, *Handbook of Mental Tests*.

<sup>2</sup> Stutsman, Rachael, *Performance Tests for Children of Pre-School Age*.

<sup>3</sup> Thomas, Dorothy Swaine and Associates, *Some New Techniques for Studying Social Behavior*, p. 122.

could not arbitrarily be considered indicative of resistance. So, also, in those tests far above the child's age level where the normal expectancy is failure, the response "I don't know" or "I can't" would usually indicate an appreciation of the difficulties and would not, without further evidence, be considered resistance. Therefore, the response "I don't know" or "I can't" was considered evidence of resistance only when the situation was unambiguous, as can be seen by a quotation from an actual record. Phyllis' record reveals a consistently resistant attitude. Peg Board B was presented and directions were given:

PHYLLIS: "Can't."

TESTER: "Of course you can."

PHYLLIS (picking up peg and placing it correctly): "I can't, I can't."

In such an instance, the response "I can't" was considered evidence of resistance.

The amount of resistance was subsequently measured by the number of times the material was presented and not accepted. For example, a given child may have resisted a certain test not once but several times before finally accepting it. Each of such incidents would count as one instance of resistance. Repeated expressions of resistance in a single presentation of test material would count as one instance of resistance.

The examiner also recorded, in those performance tests which the child was unable to complete, the length of time that the child persisted at the given task. The criterion for this test of persistence is less objective than would be ideally desired since test directions provide for encouragement, and the amount of encouragement given in each case probably varies from child to child—or even from day to day (a further reason for studying the examiner as well as the subject in future studies). However, an effort was made to encourage the child to persist to the maximum, and time was recorded only when the child definitely abandoned the project. To make the more subtle distinction between the child who per-

sists without encouragement and the child who requires considerable encouragement would have been interesting and valuable, but was impossible with the techniques here developed.

On a separate blank provided for that purpose the examiner also indicated by a check mark the test items at which occurred each instance, on the part of the child, of:

1. Gross activity—defined as actually rising from the testing table to walk or run to another part of the room.
2. Use of test materials for purposes other than that indicated by the examiner—such as building a train or tower with the colored blocks or form-boards.
3. Initiative in the test situation, subdivided into two classifications:
  - a. Spontaneously repeating test just completed.
  - b. Reaching after or asking for test material other than that presented at a given moment.
4. Each instance of spontaneous conversation—a simple remark, such as "I can't," "Where does this go?", "You tell me," "This is fun, isn't it?"—was considered one unit. Isolated words were also considered one unit, and the occasional more involved and complicated sentence was arbitrarily divided into clauses.

Spontaneous conversation was also subdivided into the following subgroupings:

- a. Expressions of satisfaction.
- b. Asking for help.
- c. Asking for praise and reassurance.
- d. "I can't."
- e. Comments on the immediate situation, such as "This is hard," "This is easy," "This is funny," etc.
- f. Collective monologue—as defined and discussed by Piaget<sup>4</sup> ("Talking to one's-self in front of others").
- g. Incidental or miscellaneous  
(f and g—subsequently recognized as unreliable—

<sup>4</sup> Piaget, Jean, *Language and Thought of the Child*, p. 18.

led to a further classification: (1) relevant or (2) irrelevant to the immediate situation).

A specially trained stenographer was present during these examinations and made verbatim reports of the total situation against which to check the tester's record. This stenographer was one who had been employed in a similar study made at the Child Development Institute of Teachers College, and the reliability of her records has been discussed by Lulu-Marie Jenkins in "A Study of the Reliability of Stenographers' and Students' Reports of Psychological Tests." Dr. Jenkins concludes:

1. The stenographers probably got 85 per cent of the entire conversation, including the test directions and practically all of the "important" incidental conversation and about half of the predetermined action in the test situation.
2. Stenographers improve in amount recorded and the accuracy of reports with practice and knowledge of the test materials, technique, and acquaintance with the examiner.
3. The errors were of a type which interfered neither with the continuity of the test nor with the picture of the test situation. Omissions of unimportant words and small parts of test directions constituted the typical error.<sup>5</sup>

Blanks similar to the one used by the tester were also filled in during these preliminary examinations by one or more persons who observed the tests. The writer is greatly indebted for this courtesy to Miss Metta M. Rust, Mrs. Mary Shattuck Fisher, Dr. Lulu-Marie Jenkins, and Dr. Dorothy Swaine Thomas.

The stenographer's reports, the check records, and the tester's records were then compared. Since there were so few cases involved in this preliminary study only a crude statistical comparison could be made. The children were ranked on the amount of each behavior item, and rank order correlations were obtained between the different observers' scores for these traits. Table 1 indicates the relationship existing between the examiner's record, the stenographer's report, and the observer's record.

<sup>5</sup> In Thomas, Dorothy Swaine, and Associates, *Some New Techniques for Studying Social Behavior*, p. 202.

TABLE I

## COMPARISON OF EXAMINER'S RECORD, STENOGRAPHER'S REPORT, AND OBSERVER'S RECORD

	RANK CORRELATION BETWEEN STENO- GRAPHIC RECORD AND EXAMINER'S RECORD	RANK CORRELATION BETWEEN OBSERVER'S RECORD AND EXAMINER'S RECORD	RANK CORRELATION BETWEEN OBSERVER'S RECORD AND EXAMINER'S RECORD
		1st Observer	2nd Observer
Spontaneous conversation (total amount) .....	.99	.99	.99
Gross activity.....	1.00	1.00	1.00
Initiative in test situation .....	1.00	1.00	1.00
Other uses for material .....	.97	.98	.98
<hr/>			
Seeking praise and reassurance ..	Stenographic report not analyzed *	.93	.78
Satisfaction expressed .....	Stenographic report not analyzed	.92	.84
Seeking help .....	Stenographic report not analyzed	.83	.67
Irrelevant conversation (9 cases)	Stenographic report not analyzed	.67	.52
Relevant conversation (9 cases) ..	Stenographic report not analyzed	.64	.49
Expression "I can't" .....	.98	.97	.98
"It's hard," "Funny," etc. ....	.98	1.00	1.00
Monologue (8 cases) .....	Stenographic report not analyzed	.87	.92
Incidental (8 cases) .....	Stenographic report not analyzed	.84	.86

\* It will be noted that the stenographic report was not analyzed in all cases. For the writer to analyze conversation where rigid definition was not possible and where the element of subjective judgment was undoubtedly an important factor would contribute little or nothing as a check on her own original record made when giving the test.

The items above the dotted line in Table I, the reliability of which was .97 or higher, plus resistance and persistence as previously defined, were the items eventually selected for study. The items below the dotted line, which for the most part are but subclassifications of conversation, were discarded.

The expressions, "I can't," "It's hard," "Funny," etc., although satisfying the criteria for high reliability were nevertheless discarded with the rest of the analysis of conversational content. Although the method used, that of rank order correlation, is recognized as being only a crude index of reliability, nevertheless, when the correlations obtained were approximately perfect it was assumed that the items studied could be considered sufficiently well defined and objective to be used as measures of behavior, without further verification. When the relationship, although positive, was less, it was recognized that the element of judgment was entering in, and such items could not have been used as indices of behavior without the continued use of an observer to determine the degree of reliability. They were therefore discarded, since the experimental work of the main study was of necessity carried by one individual. Moreover, one aim of the study was to develop a technique capable of application by the examiner herself, and simplification of technique was desirable from a purely practical standpoint. The discarded items should not, however, be omitted from all consideration, for they may possibly be most interesting and pertinent to a more complete picture of the total psychological test situation. Further refinement of method and clarification of definition would make them an admirable basis for a coöperative project.

After determining the reliability of the observer, as above described, there remained the necessity of determining the reliability of the data. These were studied by means of the odd-even technique, based on the records of the main group (obtained as discussed in Chapter III).

Before the results are presented it should be noted that odd-even numbers as they appear on the intelligence test records themselves, that is, actual numbers appearing on the test blanks, such as 1, 2, 3, 4, 5, 6, 7, 8 for the 2nd, 3rd, 4th, 5th years, etc., were not the basis for this comparison. Of necessity, not all children were given identical tests; neither was the order of presentation of the tests identical. For instance,

the range was different; the starting point was frequently different; repeated presentation of test materials, necessitated by resistance, upset any stated order; and, moreover, the order of presentation of Merrill-Palmer materials, by test direction, was determined by the children's preferences and by the exigencies of the moment. Therefore, for the purposes of recording them in this study, each presentation of test materials, regardless of the test-blank number, was given a serial number by the examiner. This was done for each child tested. Then the scores for the behavior responses studied were computed on the basis of the amount of response to the odd serial-numbered tests as compared to the total amount of response to the even serial-numbered tests. This score was computed for each child. Using the Pearson  $r$ , correlations were obtained between the odd and the even scores on the basis of the records for the ninety-one children comprising the main group.

No measure of reliability of so-called persistence is available, since no retests were given; neither could the odd-even technique be employed, since the score used was "maximum." (See Chapter V.)

Table 2 indicates the correlations obtained in the study of reliability of data. The Pearson  $r$  was used.

TABLE 2  
RELIABILITY OF DATA

	PEARSON $r$
Resistance .....	.92
Spontaneous conversation .....	.89
Spontaneous gross activity .....	.94
Spontaneously reaching or asking for test materials .....	.84
Spontaneous use of test materials for purposes other than indicated	.58 *

\* Shephard's correction for coarse grouping used in computation of last item.

The results indicate a high degree of reliability, quite sufficient to warrant the use of these measures, with the possible exception of "use of test materials for purposes other than indicated."

## CHAPTER III

### PLAN OF MAIN STUDY

The Kuhlmann Revision of the Binet Scale and the Merrill-Palmer Performance Test were administered to one hundred and two children, within three months of their third birthday—that is, chronological age was held constant within six months. The tests were given by the writer, in standardized form, in two testing periods of approximately forty-five minutes each. All but eight were given at her home in White Plains, N. Y., in a small, meagerly furnished room. Since it was found impossible to separate all children from their mothers, who brought them, each mother was asked to sit in a far corner of the testing-room, but was urged not to participate in the situation in any way. This is, of course, not an ideal arrangement, but since the mother's presence was inevitable in some cases, and since the individual reactions studied might be influenced by the mother's presence, it was thought advisable to keep the situation reasonably constant by having all mothers observe the test. The mothers as a group coöperated excellently from the standpoint of noninterference; in only four instances did they interfere to such an extent as to vitiate the test results, and the records of these tests have not been included. Three tests were incomplete—that is, the child failed to return for the second period. Another case, that of an obviously retarded Italian boy, developed into such an emotional situation that the test was abandoned. In four instances, by a subsequent checking of birthdays, the children were found to be outside the age-range. It is interesting to note that two of these errors were caused by misunderstanding of the mothers as to the age-limits imposed; two were due to falsifications of birth dates in order to secure psychological test-rating free.

For these reasons, the final analysis is based on only 91 cases; 46 of these are girls and 45 are boys. This equal division as to sex is particularly interesting in that no special effort had to be made to achieve it.

The children were obtained by writing letters to mothers of three-year-olds. The mothers were told that the Child Development Institute of Teachers College, Columbia University, was interested in studying three-year-old children in White Plains; that the study was a coöperative project, since the Institute was collecting information as to the mental and emotional development of normal children of the given age, which was important for research purposes; and that they would be given a detailed report of the psychological test findings on the individual child concerned. All examinations were, of course, free of cost.

The original list of children was obtained through the courtesy of the Board of Health of White Plains. About six months before this study, a house-to-house canvass of pre-school-age children had been made in White Plains in connection with the diphtheria antitoxin campaign. The listing made at that time was the original mailing list for this study. The letter sent to the mothers was approved by the Board of Health but did not mention the Board in any way.

As was to be expected, the response to the letters, sent out in several groups of 10 to 15 each, was meager, averaging between 7 and 8 per cent, and was obviously selective. Because of a previous arrangement with the Board of Health, the letters could not be followed up personally. However, those individuals who did respond were sufficiently interested in the project to supply a secondary mailing list of interested friends, all of whom were on the original list but had not yet been approached. Letters written to these mothers were identical with the original letter except that each was prefaced with the remark that Mrs. X had suggested that the mother addressed might be interested, etc. To some sixty-odd letters sent in this way, there were only two negative responses, and none was

ignored; whereas only 7 per cent had responded to the original letters. Moreover, fifteen such "interested friends" took the initiative and telephoned for appointments before the letters had been mailed to them. This procedure, of course, further emphasized the selectiveness of the group studied. However, selection of some sort was almost inevitable, and the claim is not made that the group studied is an unselected one. The group is defined in terms of the occupational status of the fathers, the nationality of the parents, and the I.Q.'s of the children tested. Of course it is recognized that even an "unselected sample" of this size would mean very little from the standpoint of representativeness.

The occupational level of the fathers was determined by the use of the Barr scale of occupational intelligence. This scale undoubtedly overweights the mental-test point of view with regard to the occupational hierarchy; its units are perhaps

TABLE 3  
RATINGS OF OCCUPATION LEVEL OF FATHERS

	ADULT MALE OF GENERAL POPULATION (BARR)	FATHERS OF GIFTED CHILDREN (TERMAN)	FATHERS OF THREE-YEAR- OLDS IN PRESENT STUDY
Number of cases .....	250,519	526	91
Mean rating .....	8.88	12.77	11.01
S. D. .....	3.24	2.78	3.41

<sup>1</sup> L. M. Terman, Editor, *Genetic Studies of Genius*, Vol. I, pp. 66-72.

misleading; its use most certainly does not assure complete objectivity. However, it does provide a basis of comparison of the fathers of the three-year-olds in the present study with the mean occupation rating of the entire male population of San Francisco, Los Angeles, and Oakland, as referred to by Dr. L. M. Terman, and with the mean occupation rating of the fathers of Terman's superior children.<sup>1</sup> Thus we may roughly

define, in Table 3, the occupational status of the present group in terms of the comparative data.

From Table 3 it is seen that although the group in the present study is probably definitely selected—certainly it is

TABLE 4  
OCCUPATIONS OF FATHERS OF THREE-YEAR-OLDS

OCCUPATION	NUMBER	OCCUPATION	NUMBER
Lawyer .....	10	Janitor .....	1
Salesman .....	8	Violin teacher .....	1
Stocks and bonds .....	7	Garage owner .....	1
Manufacturer (small) .....	7	Fireman .....	1
Laborer .....	6	Comptroller .....	1
Advertising executive .....	6	Chiropodist .....	1
Engineer .....	6	Sales executive .....	1
Builder .....	3	Chiropractor .....	1
Doctor .....	3	Haberdasher .....	1
Chauffeur .....	3	Navy officer .....	1
Unknown .....	3	Plumber .....	1
Traveling man .....	2	Carpenter .....	1
Accountant .....	2	Gas fitter .....	1
Train dispatcher .....	2	Gardener .....	1
Truckdriver .....	1	Writer .....	1
Mechanic .....	1	Garage man .....	1
Navy mechanic .....	1	Dishwasher .....	1

above the average of the total population—its mean rating is not so high as the mean rating of the occupations of the fathers of Terman's gifted children. Table 4 lists the occupations of the fathers of the children examined in the present study.

The racial origin of the group presents a comparatively homogeneous picture. The birthplaces of the parents and the grandparents were obtained. Both parents and all grandparents of 69 of the 91 children, or approximately 76 per cent, were born in this country. Ninety-seven per cent of the parents of the three-year-olds were born in this country; eleven per cent were Jews; two per cent were Negroes. The national background of the grandparents who were born abroad also presents a fairly homogeneous picture. They came in about

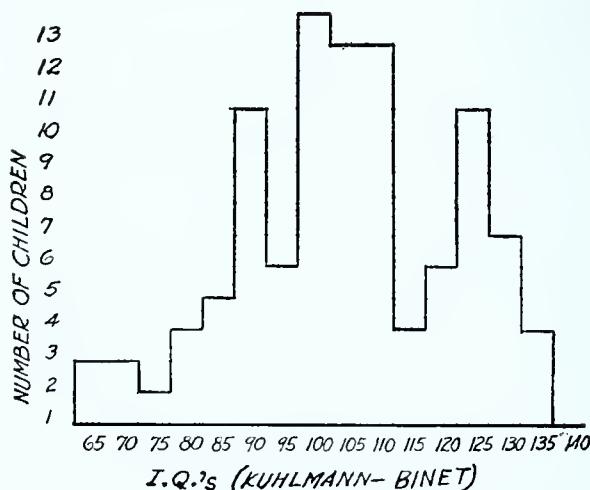


CHART I. Distribution of I.Q.'s (Kuhlmann-Binet) for 91 Three-Year-Olds

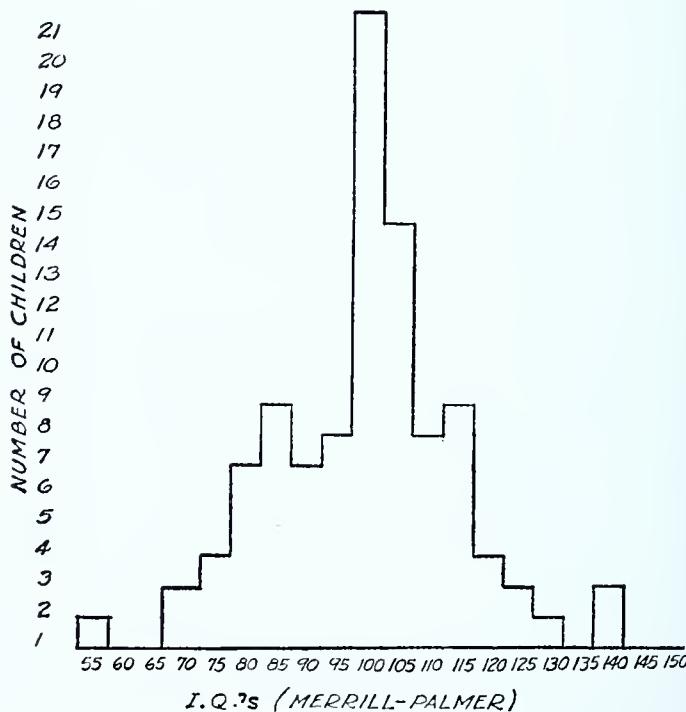


CHART II. Distribution of I.Q.'s (Merrill-Palmer) for 91 Three-Year-Olds

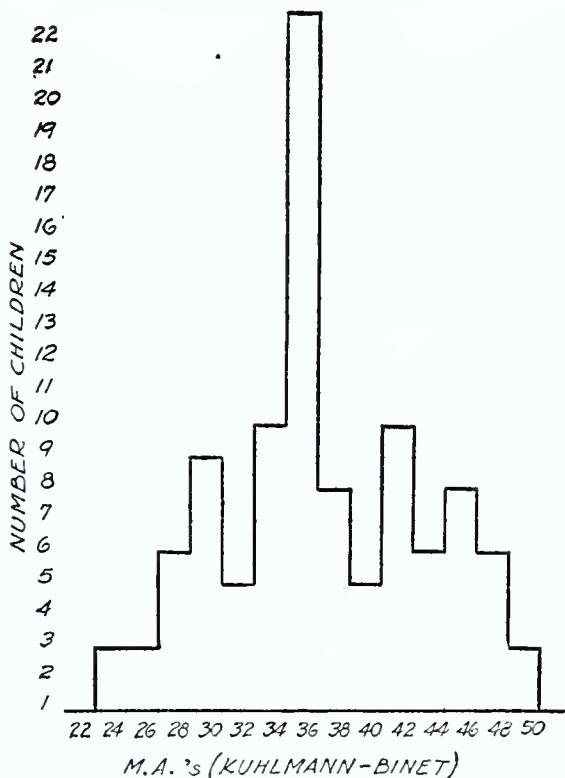


CHART III. Distribution of M.A.'s (Kuhlmann-Binet) for 91 Three-Year-Olds

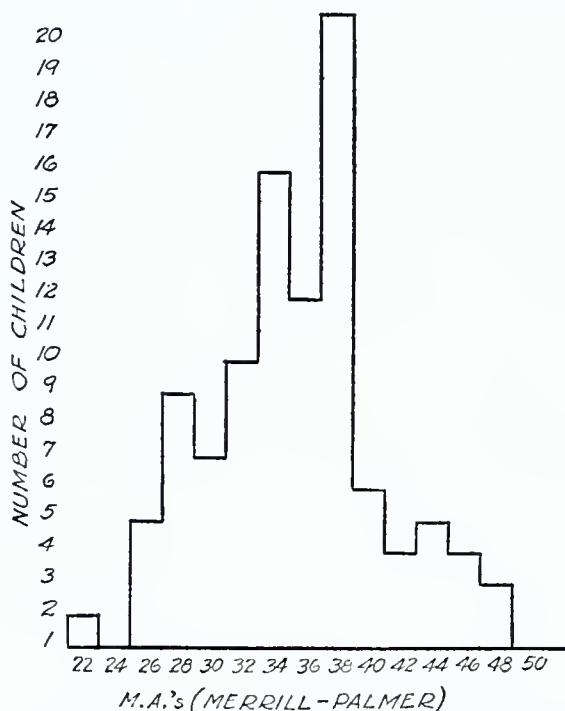


CHART IV. Distribution of M.A.'s (Merrill-Palmer) for 91 Three-Year-Olds

equal numbers from England, Scotland, Germany, and Sweden, and there were five individuals who were born in Italy.

Defining the group in terms of the intelligence of the children studied gives the following results: (it is to be remembered that the chronological age range is from 33 to 39 months). The mental age range as obtained on the Kuhlmann-Binet Scale is from 24 to 50 months; as obtained on the Merrill-Palmer Performance Scale, it is 22 to 48 months. The range in I.Q. (Kuhlmann-Binet) is from 65 to 138. The range in I.Q. (Merrill-Palmer) is from 59 to 144. The mean I.Q. of the group (Kuhlmann-Binet) is 107.6; S.D. 16.5 (Dr. Florence Goodenough, studying 100 children three years of age—chronological age range 30 to 42 months—reports a mean I.Q. of 104.4; S.D. 18.2).<sup>2</sup> The mean I.Q. (Merrill-Palmer) is 101.9; S.D. 14.8.

We may assume that the group studied is slightly above the average intelligence of three-year-olds. The absence of extremely low I.Q.'s is to be noted, and also the absence of unusually high I.Q.'s. Charts I, II, III, and IV present graphically the distribution of I.Q.'s and M.A.'s of the group.

<sup>2</sup> Goodenough, Florence, *The Kuhlmann-Binet Tests for Children of Preschool Age*, p. 40.

## CHAPTER IV

### RESISTANCE

Resistance, or refusal to accept a given test, was one of the most important reactions studied, not only because of its intrinsic interest but because of the possible comparisons with other researches in this field.

As previously noted, resistance was defined as (1) "passive resistance" by ignoring the material presented or question asked; (2) verbal resistance, such as "No," "I won't"; (3) physical resistance, manifesting itself in walking or running from the table or room, or in definitely pushing away proffered material; (4) certain instances of "I can't," or "I don't know." Each instance was recorded on the test blank opposite the test item at which it occurred, and each succeeding trial of each test item, at which resistance occurred, was also recorded. Because of the practical necessity of limiting the length of time of a mental test, and in order to make results comparable, the examiner returned to a given test item but four times. If the test was accepted on a second or third trial, as frequently happened, the maximum number of repetitions was not held necessary. When, however, a test was not accepted on the fourth presentation, it was definitely abandoned. Thus a score of four was the maximum score for resistance that could be recorded for any given test item.

This technique gives three possible methods of measuring the degree of resistance manifested by any given child. In the first place, the total amount of resistance may be recorded—that is, the number of tests resisted plus the number of times a test was repeated before it was finally accepted. The second method measures resistance by counting the number of test items resisted, with no reference to the amount of repetition

needed. The third method counts the number of tests which, regardless of repeated efforts, never were accepted. This last method lends itself, of course, to a rating of only a comparatively small proportion of the total group. The first method seems to the writer to have greatest merit if one is interested in the typical behavior reactions manifested by a given child. On the other hand, methods two and three probably affect more closely the actual test results. The third method is most important when resistance is studied in relation to its effect on the actual intelligence quotient.

Of the 91 children tested, 80 per cent finally accepted all tests; 10 per cent accepted all but 1 test; 10 per cent accepted all but from 2 to 6 tests. Interpreting these facts in terms of I.Q. it is evident that, regardless of the amount of incidental resistance offered by 80 per cent of the children, this resistance did not directly influence their intelligence test scores. Of the next 10 per cent of the three-year-olds who accepted all but 1 test, it is possible that resistance caused a variation in I.Q. of 4 points. This 4 points of possible variation is always in the direction of lowering the score. If a test is not accepted on the Kuhlmann-Binet Scale there is no alternative but to score it failed. Therefore, a possible variation in I.Q. of 4 points means that there is the chance that, had the test been accepted, the child might or might not have passed it with the chance of increasing his I.Q. 4 points. This chance is, however, reduced by the fact that it has been found that there is a disproportionately greater tendency for children to resist hard tests—that is, tests appearing on a scale at an M.A. level above their own.<sup>1</sup> The remaining 10 per cent, who refused to accept from 2 to 6 tests, run the chance of being scored from 4 to 20 points below their true I.Q.'s.

The implications of these results are encouraging to the mental tester. If they are sustained in subsequent researches, it may be concluded that resistance as evidenced in a mental

<sup>1</sup> Thomas, Dorothy Swaine, and Associates, *Some New Techniques for Studying Social Behavior*, p. 123.

test of children of 3 years of age, annoying and complicating as it is, nevertheless does not seriously impair the validity of the test rating for 90 per cent of the group. However, we must continue to be cautious lest we attach too much importance to the test rating of the remaining 10 per cent who refused ever to accept from 2 to 6 test items, and who thereby run the risk of being scored from 4 to 20 points below their true I.Q.'s. (It is assumed here that resistance, as a factor complicating the test situation, is itself extraneous to "true" intelligence.) Table 5 shows the distribution of tests never accepted and the number of children failing to accept them. Chart V presents these results graphically.

TABLE 5

DISTRIBUTION OF THREE-YEAR-OLDS FAILING TO ACCEPT  
FROM 1 TO 6 TEST ITEMS

NUMBER OF TESTS NEVER ACCEPTED	NUMBER OF CHILDREN FAILING TO ACCEPT TESTS
1 .....	9
2 .....	3
3 .....	1
4 .....	2
5 .....	2
6 .....	1

The next consideration is the number of tests resisted one or more times. Chart VI presents graphically this number in terms of the number of children resisting them. Of the 391 tests which were resisted, all but 42, or approximately 90 per cent of all tests resisted, were eventually accepted.

Resistance may also be studied from the standpoint of the type of child who is less or more resistant. For that purpose the number of tests resisted, or even the number never accepted, is probably not so important as the total amount of

resistant behavior exhibited during the total test procedure. Therefore, resistance is here measured in terms of the number of tests resisted plus the number of times each resisted test

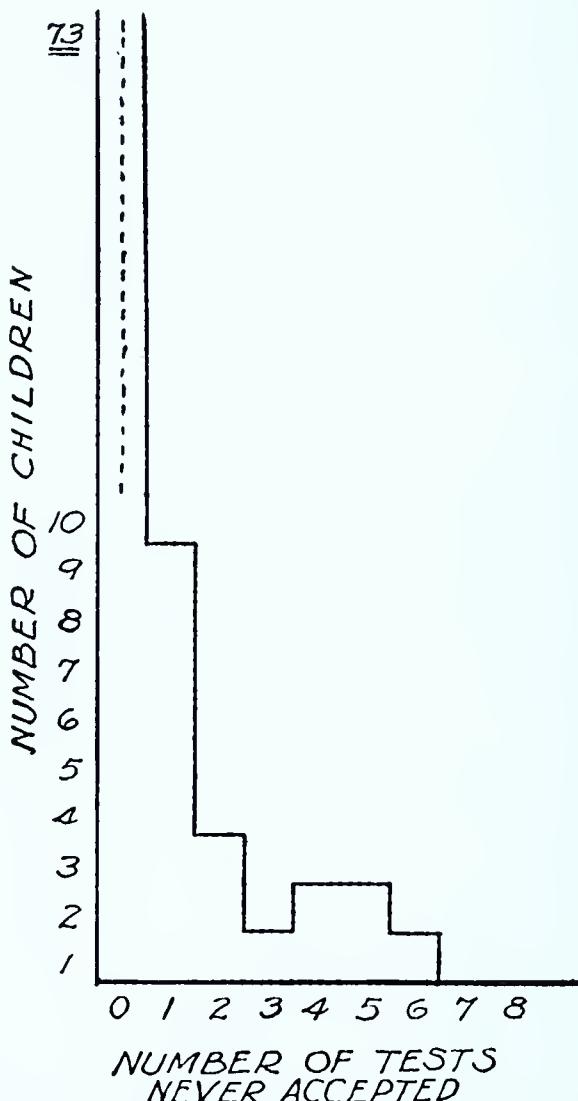


CHART V. Number of Tests Never Accepted in Terms of Number of Children Failing to Accept Them

was presented again and resisted, whether it were one, two, three, or four times. Chart VII gives a graphic representation of the total amount of resistance in terms of the individual

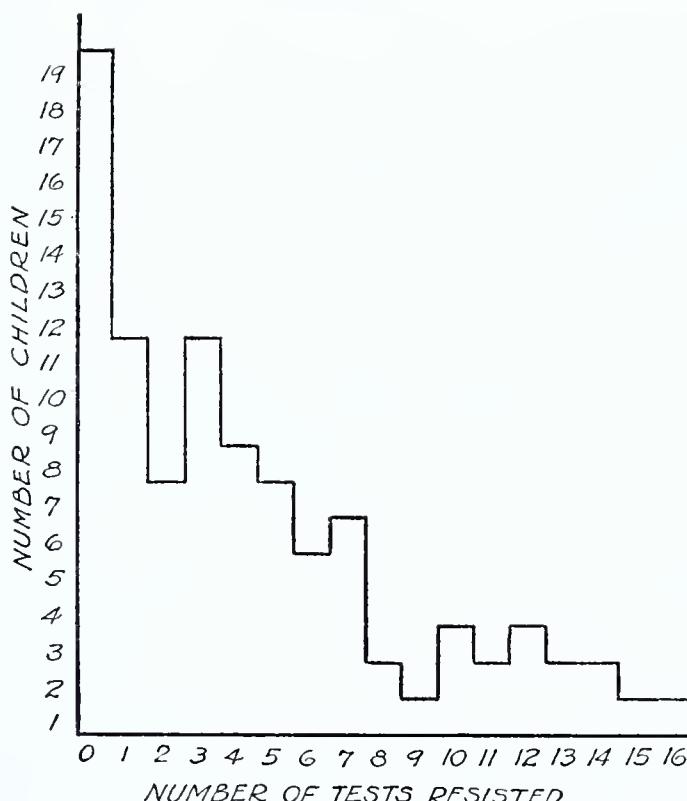


CHART VI. Distribution of Number of Tests Resisted in Terms of Number of Children Resisting Them

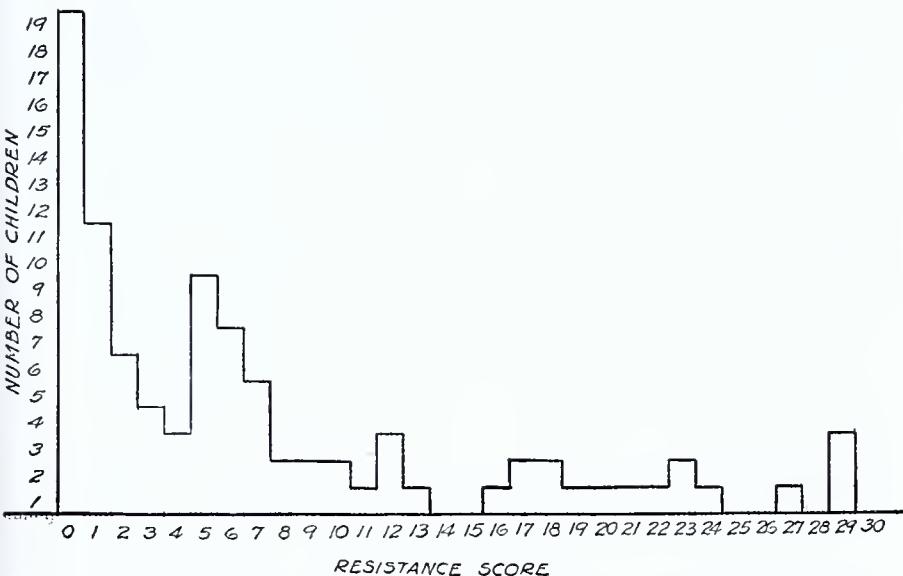


Chart VII. Distribution of Total Resistance Scores of 91 Three-Year-Olds

children tested. The range in resistance thus measured is from 0 to 29; the mean is 7.2; the S.D. is 7.98.

Resistance, thus measured, was studied in relation to the intelligence rating of the children. Table 6 shows the relation of Kuhlmann-Binet M.A. and I.Q. to resistance, the group being divided above and below the median of resistance. The middle score is 5.

TABLE 6

MEAN I.Q. AND M.A. (KUHLMANN-BINET) OF GROUPS FALLING ABOVE AND BELOW MEDIAN RESISTANCE SCORE

	KUHLMANN-BINET MEAN M.A.	S.D.	KUHLMANN-BINET MEAN I.Q.	S.D.
Below median of resistance .....	40.5	6.9	111	17
Above median of resistance .....	36.5	6.0	103	16

The difference between the mean I.Q.'s is  $8 \pm 2.4$ . The difference between the mean M.A.'s is  $4 \pm .96$ . In the latter case the difference between the means is at least four times the probable error, and, although slight, the difference may probably be considered significant. The difference between the means of the I.Q.'s is not four times its probable error though it approaches that amount. One must remember, in considering the statistical results that are presented in this study, that the smallness of the group, plus its relative homogeneity, will tend to minimize any statistical differences. While chronological age has been kept constant within six months, nevertheless, judging from results of previous researches, variations within that limit probably contribute to the difference in results obtained from considering I.Q. as compared with M.A.

It is interesting to compare the results of Table 6 with the very crude results obtained by the writer in a previous study of 28 nursery-school children with a chronological age-range of from 18 to 48 months.<sup>2</sup> That the group was more definitely

<sup>2</sup> Thomas, Dorothy Swaine, and Associates, *Some New Techniques for Studying Social Behavior*, p. 133.

superior than the present one studied is quite obvious from a consideration of the mental age range, which was 30-72 months. However, resistance to the same tests was measured in the same way. The results of these measurements are shown in Table 7.

The statistical significance of these differences was not obtained, as the number of cases was too small. However, in so

TABLE 7

MEAN CHRONOLOGICAL AGE, M.A. AND I.Q. OF GROUPS FALLING ABOVE AND BELOW MEDIAN OF RESISTANCE

	MEAN C.A.	MEAN M.A.	MEAN I.Q.
Below median of resistance .....	41.9	62.7	134.0
Above median of resistance .....	37.7	43.2	113.5

far as differences existed between the mean I.Q. and the M.A. ratings of the two groups, they were considered indications of possible tendencies for the older, brighter group to be, on the average, less resistant.

Both these studies may profitably be compared with that of Dr. Martha May Reynolds on the negativism of preschool children.<sup>3</sup> Dr. Reynolds studied 229 children, ranging in chronological age from 1 year 9 months to 5 years 6 months, in ten situations selected with a view to evaluating resistant behavior. The situations included surrender of personal liberty (tester picked child up in lap), imitation of movements (clapping hands, etc.), "don't" situations (after child had been playing with blocks he was told, "Don't make that any more, John"), repetition of one and two digit series, reestablishment of *rappoport* after a neglect period, etc., etc. These children were given the Kuhlmann-Binet test, also. Among other interesting results, Dr. Reynolds found that the coefficient of

<sup>3</sup> Reynolds, Martha May, *Negativism of Preschool Children*.

correlation between chronological age and negativism was  $-.53 \pm .03$ ; that the coefficient of correlation between M. A. and negativism was  $-.48 \pm .03$ ; and that the coefficient of correlation between I.Q. and negativism was  $-.09 \pm .04$ .<sup>4</sup>

Coefficients of correlation between resistance and chronological age, M.A., and I.Q., based on the present study of 91 three-year-olds, yield the results indicated in Table 8.

TABLE 8  
CORRELATION BETWEEN RESISTANCE AND CHRONOLOGICAL AGE, MENTAL AGE,  
INTELLIGENCE QUOTIENT (KUHLMANN-BINET)

VARIABLE	Pearson <i>r</i>	P.E.
<i>r</i> between C.A. and resistance .....	-.14	.07
<i>r</i> between M.A. and resistance .....	-.32	.06
<i>r</i> between I.Q. and resistance .....	-.29	.06

Blakeman's test for linearity was used with the following results:<sup>5</sup>

Resistance and chronological age .....	2.736 < 4.047
Resistance and mental age .....	3.774 < 4.047
Resistance and intelligence quotient .....	3.584 < 4.047

One does not feel certain, however, that the coefficient of correlation can justifiably be considered a valid measure of relationship between factors shown in Table 8, since in no case, except possibly in connection with the chronological age, was the left-hand side of the equation perceptibly smaller than 4.047. Therefore, results obtained by using the correlation ratio are presented in Table 9. Scatterdiagrams of resistance scores and C.A., M.A., and I.Q. are also presented in Charts VIII, IX, and X.

It is interesting to note that, in Tables 7 and 9, the degree of relationship between chronological age and resist-

<sup>4</sup> *O.p. cit.*, pp. 120-21.

<sup>5</sup> Holzinger, Karl J., *Statistical Methods for Students in Education*, p. 183.

ance is very small, especially as compared with Dr. Reynolds' results. Since, however, the age range in the present study was limited to six months, a perceptible decrease in the amount of relationship might be expected. Similarly, because of the narrow age range of the group studied, the relationship between M.A. and resistance, as expressed by the Pearson  $r$  ( $r = -.32 \pm .06$ ), is small in comparison with Dr. Reynolds'  $r$ , which is  $-.48 \pm .03$  (if coefficients of correlations may be compared; strictly speaking, they may not). A definite negative relationship exists between resistance and I.Q., as may be observed in Table 9. The fact that this result does not coincide with the very slight correlation reported by Dr.

TABLE 9

CORRELATION RATIOS BETWEEN RESISTANCE AND CHRONOLOGICAL AGE, MENTAL AGE, INTELLIGENCE QUOTIENT (KUHLMANN-BINET)

VARIABLE	$N_{xy}$	$N_{yz}$
Chronological age and resistance .....	— .18	— .32
Mental age and resistance .....	— .43	— .51
Intelligence quotient .....	— .55	— .48

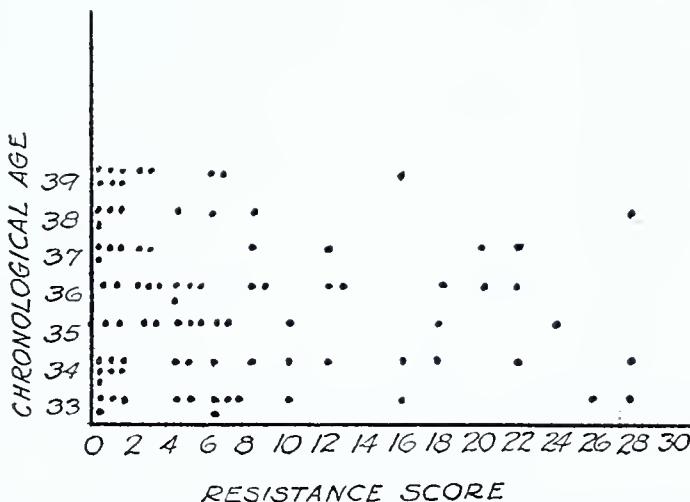


CHART VIII. Scatterdiagram Showing Relationship Between Chronological Age and Resistance

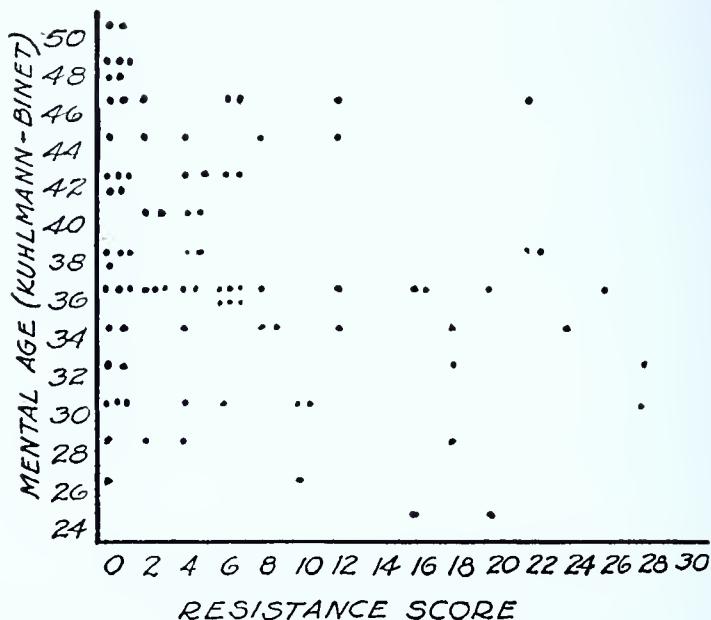


CHART IX. Scatterdiagram Showing Relationship Between M. A. (Kuhlmann-Binet) and Resistance

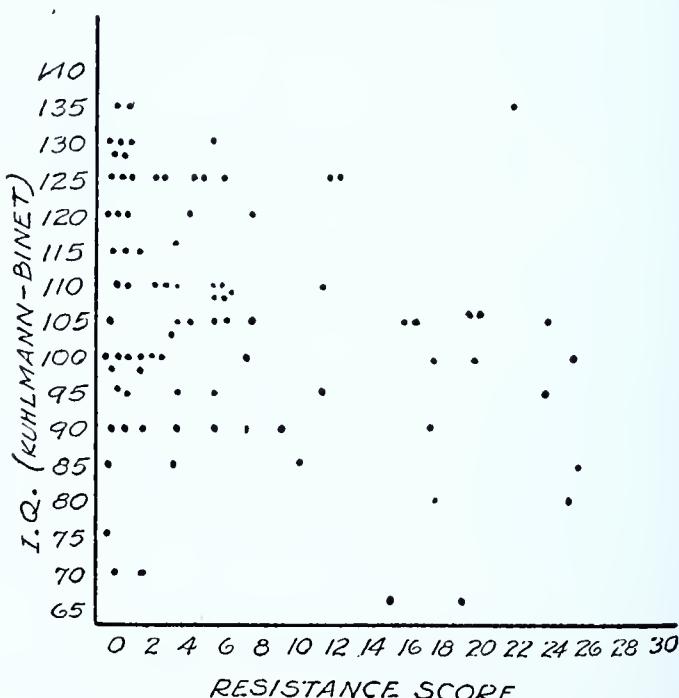


CHART X. Scatterdiagram Showing Relationship Between I.Q. (Kuhlmann-Binet) and Resistance

Reynolds may be due to the effect of the limited age range in the present study where differences in M.A. and I.Q. would probably be negligible.

Because of the set-up of the study, no conclusion can be reached concerning the relationship between chronological age and resistance. The study lends itself, however, to a verification of the fact that at least in children three years of age, less, rather than more, resistance tends to be associated with a high M.A.

It is interesting at this point to compare the ratings of the more and the less resistant groups as obtained on the Kuhlmann-Binet Scale with the ratings obtained on the Merrill-Palmer Performance Scale. Table 10A present these results.

TABLE 10A

MEAN M.A. AND I.Q. ON KUHLMANN-BINET AND MERRILL-PALMER SCALES OF GROUPS FALLING ABOVE AND BELOW MEDIAN RESISTANCE SCORE

	ABOVE MEDIAN	BETWEEN MEDIAN	DIFFERENCE	P.E. OF DIFFERENCE
Mean I.Q. (Kuhlmann-Binet) .	103.0	111.0	8.0	2.4
Mean I.Q. (Merrill-Palmer) ...	100.5	103.0	3.5	2.1
Mean M.A. (Kuhlmann-Binet)	36.5	40.5	4.0	.96
Mean M.A. (Merrill-Palmer) .	35.6	37.9	2.3	.85

TABLE 10B

MEAN M.A. AND I.Q. OF GROUP FALLING ABOVE AND BELOW MEAN SCORE OF RESISTANCE

	ABOVE MEAN	BETWEEN MEAN	DIFFERENCE	P.E. OF DIFFERENCE
Mean I.Q. (Kuhlmann-Binet) .	100.3	111.0	10.7	2.6
Mean I.Q. (Merrill-Palmer) ...	99.1	103.2	4.1	2.4
Mean M.A. (Kuhlmann-Binet)	35.3	39.8	4.5	.9
Mean M.A. (Merrill-Palmer) .	35.2	37.4	2.2	.9

Because of the skewness of the distribution of the resistance scores, the groups were again divided above and below the *mean* resistance score. Table 10B presents these results.

Comparing Tables 10A and 10B, we find but little difference. The differences between the mean I.Q.'s of the more and the less resistant groups are perhaps slightly emphasized when the total group is divided at the mean resistance score rather than at the median.

The difference between the M.A.'s on the Kuhlmann-Binet test is  $4.0 \pm .96$ ; the difference between the M.A.'s on the Merrill-Palmer test is  $2.3 \pm .85$ . Both differences are in the same direction and imply that the less resistant group has, on the average, the higher mental age. But the difference on the Kuhlmann-Binet test is larger than that on the Merrill-Palmer test; it may, moreover, be considered statistically significant; whereas the difference in mean ratings obtained on the Merrill-Palmer test is not. A partial explanation of this condition is undoubtedly furnished by the fact that the Merrill-Palmer test includes a method of scoring resisted responses instead of counting them out-and-out failures. Unfortunately Miss Rachael Stutsman, the originator of the Merrill-Palmer Scale, has not yet published her statistical method of arriving at her scoring device, and, therefore, further discussion would be futile. It should simply be noted that a greater average difference between the more and the less resistant groups exists when M.A. is measured on the Kuhlmann-Binet Scale than when it is measured on the Merrill-Palmer Scale.

## CHAPTER V

### PERSISTENCE

Persistence, as defined in this study, is measured by the time spent at a form-board without obtaining a successful solution within a given time-limit.

This chapter is based on 85 cases, instead of 91, because in six cases all form-boards included in the test scale were completed and no persistence score could be computed for them. Persistence was measured on the following test items in the Merrill-Palmer Scale: Peg Boards, Buttons, Sequin Form-Board, and the Mare and Foal. The Manikin, Decroly Matching Pictures, and the Pyramids were not included, since mistakes are not necessarily obvious to the child, and it is possible that intelligence in recognizing mistakes, rather than persistence, might motivate continuance of the project.

In some instances more than one measure of persistence could be obtained (e.g., Peg Boards and Buttons) but, since this condition was not uniform, one instance only has been used. When more than one response was obtained, the highest was arbitrarily chosen as the score. Moreover, since the tests provided no scoring basis for time limits above 371 seconds (and this only for the Mare and Foal) and because of the practical necessity of limiting the testing period, the test item was arbitrarily abandoned after six and a half minutes unless there were reasons for continuing it, peculiar to the child's individual psychology. This condition doubtless explains the lumping of scores at 375 to 400 as shown in Chart XI, for in no instance was a score higher than 390 given. It might not be out of order to reiterate here that this study attempts to measure behavior responses only as they are discernible and can be measured within the practical limits of a mental test.

A source of error occasionally occurs when a child so responds to a difficult situation that, unless *rapport* be entirely lost, the examiner takes the initiative in abandoning a test before six and a half minutes have elapsed. Fortunately this occurred but twice in this experiment, and, fortunately again, another persistence score was available in each case. Incidentally, this characteristic in itself, if we could be assured of the sound judgment of the examiner, is a valuable index of a reaction which is an important and significant individual difference; but since it remains in the realm of judgment and impression it has not been considered in this study. So far as the writer is aware, there has been no other research in the field of preschool psychology dealing with sustained effort. Therefore the results of this study cannot be corroborated from other sources.

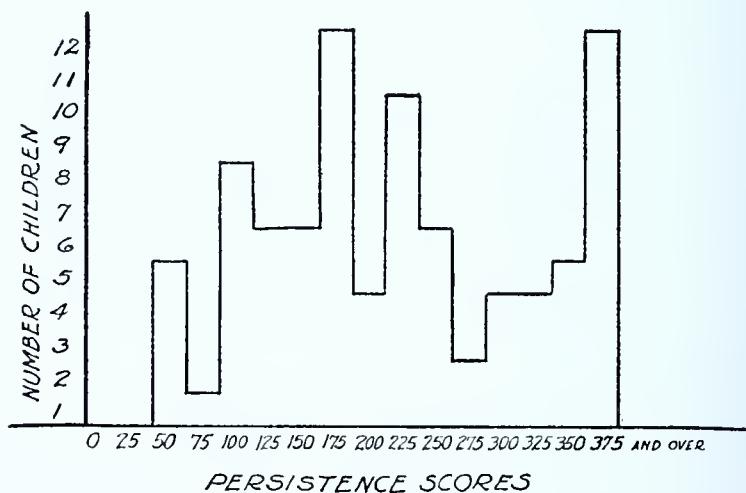


CHART XI. Distribution of Persistence Scores of 85 Three-Year-Olds

Chart XI shows the distribution of persistence scores. The range is 57 seconds to 390 seconds or above. The mean is 236.4; the S.D. is 107.6. The median is 226. The groups above and below this median were studied in terms of I.Q. and M.A. on both intelligence tests. Table 11 presents these results.

The greatest difference in mean scores occurs when they are considered in terms of mean I.Q.'s on the Kuhlmann-Binet

TABLE II

## MEAN I.Q. AND M.A. OF GROUPS FALLING ABOVE AND BELOW MEDIAN SCORE OF PERSISTENCE

	ABOVE MEDIAN	BETWEEN MEDIAN	DIFFERENCE	P.E. OF DIFFERENCE
Mean I.Q. (Kuhlmann-Binet) .	109.6	104.5	.5.1	2.4
Mean I.Q. (Merrill-Palmer) ..	100.9	101.1	.2 *	2.0
Mean M.A. (Merrill-Palmer) .	38.6	36.8	.1.8	.9
Mean M.A. (Kulhmann-Binet)	36.7	36.4	.3	.3

\* Difference of .2 is in opposite direction, *i.e.*, below median group is .2 higher.

test; that is, there is a slight tendency for the more persistent group to have a higher mean intelligence score. However, since the difference is but approximately twice its probable error, it cannot be considered statistically significant. The other differences are negligible. This fact is especially interesting when it is remembered that persistence, as defined here, was measured by the response to test items on the Merrill-Palmer Scale, and that, therefore, such differences as might exist, might reasonably have been expected to show up more clearly in results obtained in terms of Merrill-Palmer ratings. They did not do so.

No computations have been made on the basis of the mean score as the dividing point. Inspection of the distribution of scores (Chart XI) shows that, because of the lumping of scores at 375, due to the arbitrary discontinuance of the test at 390 seconds, the mean score would not be a reliable basis for the division of groups. The median, of course, is not thus affected.

Interpreting the results, from a consideration of both intelligence quotient and mental age, we must conclude that there is no significant relationship between persistence as it has been measured, and intelligence as measured by the mental tests. However, it is possible that, were a less selected group studied, a higher degree of intelligence or mental age

might be associated with greater persistence—that is, the homogeneity and smallness of the group may be obscuring small differences. However, we have learned that three-year-olds do differ in this measure of persistence. A crucial measure of the significance of this trait would of course lie in terms of achievement, as contrasted with potential achievement which, theoretically at least, the mental tests measure. Since the present study was limited to the intelligence test situation, it gives no basis for this further study.

Another interesting comment can be made concerning the relative amount of persistence on the button test as compared with the persistence on the form-boards. A study of this point was prompted by repeated observation of children's attitudes toward the Decroly buttons. It had been previously found that more than average resistance was manifested toward the button test,<sup>1</sup> and it was the writer's impression that, even when the test was accepted, resistance, or at least unwillingness to do it, carried over and possibly could be measured by the smaller time interval the child persisted at the task.

In studying the results of the present study it was found that exactly half the total number of children persisted less on the buttons than on other material presented. There was one case where the buttons were consistently refused and never accepted, one case where, for some reason, the test was entirely omitted, and eighteen cases where all three button tests were completed, so that no persistence score was available for them. That is, of the 85 tests for which a maximum persistence score was available, there were only 64 cases where a persistence score was also available on the button test. Of these, 43, or 67 per cent of the cases, showed less sustained effort on the button test than on the form-boards.

These results are not very convincing. From the standpoint of methodology, however, they are interesting. There is at present in the field of social case-work, a tendency to consider

<sup>1</sup> Thomas, Dorothy Swaine, and Associates, *Some New Techniques for Studying Social Behavior*, p. 125.

results obtained in individual studies capable of expansion and general application without further verification. As a basis for further objective research such results are undoubtedly to be welcomed, and, as in the case of the clinical approach to behavior where there is no adequate measure of many of the attitudes and relationships involved, our "hunches," evolved from experience and repeated observation, may be better than nothing, but it is essential that these opinions be reduced to a more objective basis, one that will permit of the rating both of the relative degree of divergence from established norms and of the evaluation of the seriousness of such divergence. The writer is frank to admit that, from her own experience and from that of several of her associates, there has grown up quite a definite impression of inadequacy regarding the button test, because of the children's comparatively slight interest in it. The results just quoted support that impression to a slight degree, but inevitably introduce the need of caution in the interpretation of behavior based only on uncontrolled observation, and indicate the selectiveness of the basis of many impressions.

## CHAPTER VI

### SPONTANEOUS CONVERSATION

It was noted in the chapter on method that an attempt had been made to classify the spontaneous conversation of the child into meaningful categories. It was also noted that this attempt had been abandoned when reliability coefficients, while indicating a comparatively high degree of relationship between different observers' records of the same situation, nevertheless showed considerable variation. This was interpreted as indicating that the element of judgment was confusing the records and that, therefore, such classifications could not be considered the basis of any real measurement. We were able, however, to obtain practically perfect agreement as to the *amount* of spontaneous conversation, and it is in these terms that our results have been studied.

Disregarding any influence that the factor of spontaneous conversation may have had on the mental test rating as such, we found that, as was the case with the other factors so far studied, the children's scores spread widely on a distribution curve in regard to this trait. That is, quite apart from any implications as to its relationship with intelligence, a measure of the amount of spontaneous conversation discriminates among a given group of children.

It is tempting, of course, to interpret such differences speculatively. It is possible that overt behavior as measured in terms of spontaneous conversation and activity (the subject of the next chapter) may be indicating, to use the clinical expression, an extraverted type as contrasted with an introverted type of child. Since we have no known measure of such characteristics of children at a preschool level, we have no scientific basis for comparison or for the definite establishment

of any such assumption. The idea must remain in the realm of interesting possibilities and await the development of other techniques to prove or disprove our suggestion. Nevertheless, if spontaneous expression in terms of conversation and activity is important in our consideration of a given child, we have here a means of evaluating this expression—obtained in a specific and relatively controlled situation.

As previously noted, a score of one may indicate a single clause or phrase as well as the isolated single words which children spontaneously use. In the present study the range is from 0 to 386. Chart XII indicates the distribution obtained, in step intervals of 25. The mean is 141.3; the S.D. is 95.3. As in the method used in studying resistance and persistence, the group was divided above and below the middle score, which is 126, and studied in terms of mean I.Q. and M.A. on the Kuhlmann-Binet and Merrill-Palmer tests. Table 12A shows these results.

TABLE 12A

MEAN I.Q. AND M.A. OF GROUP ABOVE AND BELOW MEDIAN SCORE OF SPONTANEOUS CONVERSATION

	GROUP ABOVE MEDIAN	GROUP BELOW MEDIAN	DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	113.1	102.0	11.1	2.3
Mean I.Q. (Merrill-Palmer) .....	103.8	100.1	3.1	2.0
Mean M.A. (Kuhlmann-Binet) .....	39.7	37.0	2.7	.9
Mean M.A. (Merrill-Palmer) .....	37.4	36.3	1.1	.7

It is interesting to note that there seems to be a statistically significant difference between the I.Q.'s on the Kuhlmann-Binet Scale of the two groups. This difference becomes less when considered in terms of M.A., probably indicating the influence of chronological age even though that was held constant within six months. The difference becomes quite negligible when viewed in terms of I.Q. and M.A. as obtained on the

Merrill-Palmer Scale. In interpreting this fact, it must be remembered that the Kuhlmann-Binet Scale weights the language factor considerably more than the Merrill-Palmer Scale does. Therefore, language facility would be expected to earn

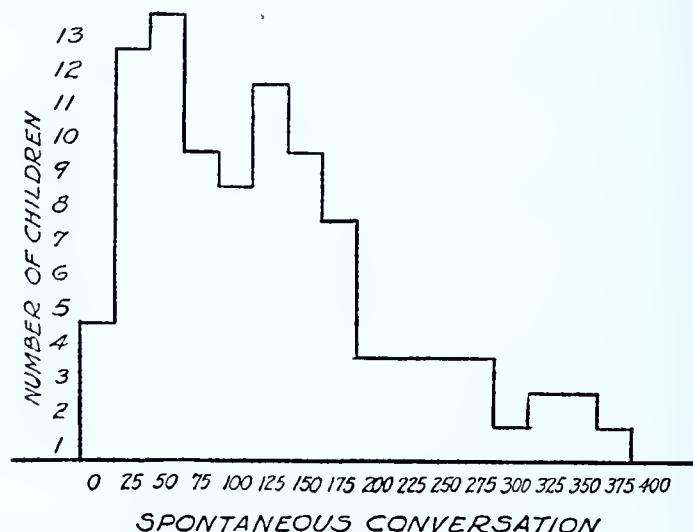


CHART XII. Distribution of Scores in Spontaneous Conversation of 91 Three-Year-Olds

for a given child a higher rating on the Kuhlmann-Binet test than on the Merrill-Palmer test. If the language facility, as rated by the tests, is correlated with amount (and possibly ease) of language expression, we would expect to find the results indicated above.

TABLE 12B

MEAN I.Q. AND M.A. OF GROUPS FALLING ABOVE AND BELOW MEAN SCORE OF SPONTANEOUS CONVERSATION

	ABOVE MEAN	BELOW MEAN	DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	113.2	102.5	10.7	2.2
Mean I.Q. (Merrill-Palmer) .....	104.7	99.9	4.8	2.1
Mean M.A. (Kuhlmann-Binet) .....	39.9	37.1	2.8	.8
Mean M.A. (Merrill-Palmer) .....	37.3	36.1	1.2	.8

Table 12B presents the results, in terms of M.A. and I.Q., when the group was divided on the basis of the mean score in spontaneous conversation.

It will be noted that the results shown in Table 12B are practically the same as those of Table 12A, where the groups were divided on the basis of the median score.

A factor which should probably be taken into account, however, is the number of test items to which a child was exposed. The number is, of course, not constant in all cases, depending as it does on the range of a given child's abilities. The brighter child usually takes more tests. It was thought that possibly this factor was influencing our results. Incidentally, the need for any such correction was not very clearly indicated in the study of resistance, since continued resistance itself tended to increase the number of test items given. In the study of persistence a need for such a correction was not indicated in any way.

Table 13 indicates the results of computing mean differences of the groups above and below the median of spontaneous conversation when, instead of using the quantitative amount of conversation outright, it was in each case divided by the number of tests.

TABLE 13

MEAN I.Q. AND M.A. OF GROUPS ABOVE AND BELOW MEDIAN SCORE OBTAINED BY DIVIDING AMOUNT OF SPONTANEOUS CONVERSATION BY NUMBER OF TESTS

	ABOVE MEDIAN	BETWEEN MEDIAN	DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	112.5	103.6	8.9	2.3
Mean I.Q. (Merrill-Palmer) .....	103.5	100.2	3.3	2.0
Mean M.A. (Kuhlmann-Binet) .....	39.4	37.6	1.8	.8
Mean M.A. (Merrill-Palmer) .....	37.0	36.2	.8	.7

Comparing Tables 12A and 13, we see that the differences in Table 13 are in the same direction as those in Table

12A, but that dividing the amount of spontaneous conversation by the number of tests has reduced differences until they should probably be considered negligible, except possibly for the difference in the mean I.Q.'s as obtained on the Kuhlmann-Binet Scale.

## CHAPTER VII

### SPONTANEOUS ACTIVITY (GROSS)

Spontaneous activity (gross) was defined as a child's actual movement about the room, that is, leaving the testing chair and table, whether for the purpose of running to look out of the window, actually leaving the room, or spontaneously getting new or old test materials, etc. Each such instance was considered a unit and given a score of one.

The range in amount of activity is from 0 to 60. The mean is 11.1; the S.D. is 12.1. Chart XIII shows the distribution, in step intervals of 2. It is readily to be observed that the distribution is not that of a normal distribution curve.

By dividing the groups above and below the median point, which falls at 8, the mean I.Q.'s and M.A.'s were computed. Table 14A gives these results.

TABLE 14A  
MEAN I.Q. AND M.A. OF GROUPS FALLING ABOVE AND BELOW MEDIAN SCORE  
OF GROSS ACTIVITY

	ABOVE MEDIAN	BELOW MEDIAN	MEAN DIFFER- ENCE	P.E. OF THE DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	110.1	104.4	5.7	2.4
Mean I.Q. (Merrill-Palmer) .....	103.6	100.0	3.6	2.1
Mean M.A. (Kuhlmann-Binet) .....	38.8	37.7	1.1	.9
Mean M.A. (Merrill-Palmer) .....	36.9	36.9	0	.8

It will be noted that no statistically significant differences emerge from Table 14A. The slight differences which are observable indicate a tendency for the more active group to

obtain slightly higher intelligence scores on the Kuhlmann-Binet Scale. Differences in mean ratings on the Merrill-Palmer Scale are even less significant.

Because of the skewness of the distribution, the groups were also divided on the basis of the mean score for spontaneous activity. Table 14B presents the results.

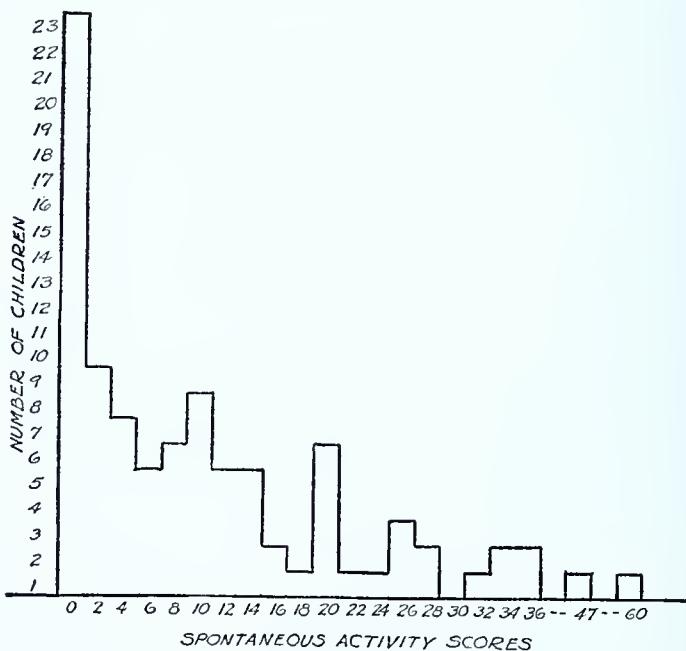


CHART XIII. Distribution of Scores in Spontaneous Activity of 91 Three-Year-Olds

Despite the skewness of the distribution, these results do not differ markedly from those of Table 14A, where the median score divided the groups. In fact, the differences which do exist are absolutely negligible.

As in studying amount of spontaneous conversation, it was thought that length of time of the testing period, as indicated by the number of tests given, might alter the results. Therefore, the amount of spontaneous gross activity was divided by the number of test items. It was found that when the group was divided above and below the median point of such a score, although there was some slight shifting of rank, the cases

TABLE 14B

MEAN I.Q. AND M.A. OF GROUPS FALLING ABOVE AND BELOW MEAN GROSS ACTIVITY SCORE

	ABOVE MEAN	BELOW MEAN	DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	110.2	105.3	4.9	2.3
Mean I.Q. (Merrill-Palmer) .....	102.1	101.8	.3	1.4
Mean M.A. (Kuhlmann-Binet) .....	38.6	37.9	.7	.9
Mean M.A. (Merrill-Palmer) .....	36.1	36.9	.8 *	.8

\* Difference in opposite direction, *i.e.* group below mean, higher.

remained in two groups, identical with the grouping when only activity was counted. Therefore, the mean I.Q. and M.A. ratings for the two groups remain as previously noted in Table 14A.

## CHAPTER VIII

### "INITIATIVE" INTEREST IN TEST MATERIALS

The title of this chapter demands explanation. The overt behavior items measured were instances where the child spontaneously expressed initiative in the test procedure. The use of a blanket term, such as "initiative," would probably be assuming a validity and breadth of connotation which is not justified. Therefore, the type of behavior measured is indicated simply by specific definition. One unit, or a score of one, was given when: (1) the child spontaneously repeated or attempted to repeat a given test without direction from the examiner; (2) when the child reached, or asked for, materials other than those present at a given moment. (For instance, the test materials, especially the Merrill-Palmer materials, in their attractive boxes, were on a rack by the examiner's side, and it frequently happened that a child expressed interest in a box, the contents of which had been previously disclosed, or in the balls or pictures, or in "that other box.")

Chart XIV indicates the distribution of scores obtained by such measurements. The range is from 0 to 43. The mean is 16.0, and the standard deviation is 9.5.

As in the analysis of the other behavior responses, the group

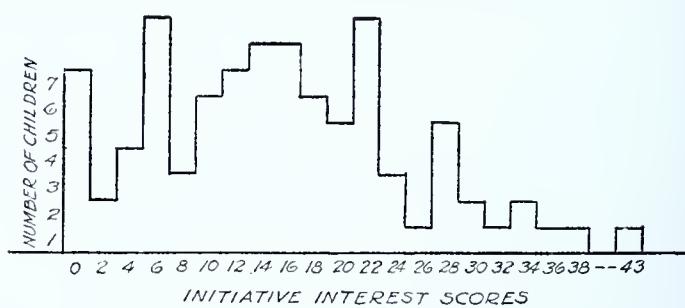


CHART XIV. Distribution of Scores in Initiative Interest in Test Materials of 91 Three-Year-Olds

was divided above and below the median score, which in this case falls at 15. Table 15A shows the means of the two groups in terms of I.Q. and M.A. on the two scales, as well as the mean difference between the two groups.

TABLE 15A

MEAN I.Q. AND M.A. OF GROUPS FALLING ABOVE AND BELOW MEDIAN POINT OF INITIATIVE INTEREST IN TEST MATERIALS

	ABOVE MEDIAN	BETWEEN MEDIAN	DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	107.6	106.5	1.1	2.5
Mean I.Q. (Merrill-Palmer) .....	103.2	100.5	2.7	1.3
Mean M.A. (Kuhlmann-Binet) .....	38.9	38.0	.9	.9
Mean M.A. (Merrill-Palmer) .....	37.5	35.6	1.9	.8

Inspection of Table 15A reveals that no statistically significant differences exist. There is a very slight tendency for the group above the median to have a higher mean intelligence test score than the group falling below the median, a difference slightly larger when measured in terms of M.A. and I.Q. ratings on the Merrill-Palmer test than when measured in terms of M.A. or I.Q. on the Kuhlmann-Binet test.

TABLE 15B

MEAN I.Q. AND M.A. OF GROUPS FALLING ABOVE AND BELOW MEAN SCORE OF INITIATIVE INTEREST IN TEST MATERIALS

	ABOVE MEAN	BETWEEN MEAN	DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	109.1	107.8	1.3	2.5
Mean I.Q. (Merrill-Palmer) .....	102.6	100.5	2.1	1.3
Mean M.A. (Kuhlmann-Binet) .....	38.3	38.4	.1 *	.9
Mean M.A. (Merrill-Palmer) .....	37.0	35.9	1.1	.7

\* Difference in opposite direction—*i.e.* group below mean, higher.

The group was also divided on the basis of *mean score*. Table 15B presents the results.

Comparing the results shown in Table 15B with those in Table 15A, where the median score was the basis of the division of the group, we find only negligible differences.

## CHAPTER IX

### SPONTANEOUS USE OF TEST MATERIALS

A marked difference exists among children in their interest in using test materials, such as blocks and form-boards, for purposes other than those indicated in a given test procedure. The spontaneous building of "choo-choo trains" and bridges with the colored blocks is not an unusual procedure.

Each instance of such an activity in connection with a single test item in the present study was scored one. Chart XV shows the distribution of scores obtained by such measurements. The range is from 0 to 10; the mean is 2.4; the standard deviation, 2.16.

The group was divided above and below the median score, which falls at 2, and studied in terms of I.Q. and M.A. Table 16A indicates these results.

No statistically significant differences emerge from the analysis shown in Table 16A, although such differences as do exist indicate a tendency for the group falling above the median to have slightly higher intelligence ratings than the group falling below the median.

TABLE 16A

MEAN I.Q. AND M.A. OF GROUPS FALLING ABOVE AND BELOW MEDIAN SCORE OF  
"USE OF TEST MATERIALS FOR PURPOSES OTHER THAN THOSE INDICATED"

	ABOVE MEDIAN	BELOW MEDIAN	DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	110.6	105.6	5.0	2.3
Mean I.Q. (Merrill-Palmer) .....	104.4	99.6	4.8	2.0
Mean M.A. (Kuhlmann-Binet) .....	38.1	37.5	.6	.9
Mean M.A. (Merrill-Palmer) .....	37.8	35.4	2.4	.8

However, it was thought that important differences might have been obscured by the fact that 18 cases fell at the median. Therefore, mean ratings were obtained for the groups falling above and below the median score, omitting these 18 cases from the computations. Table 16B presents these results.

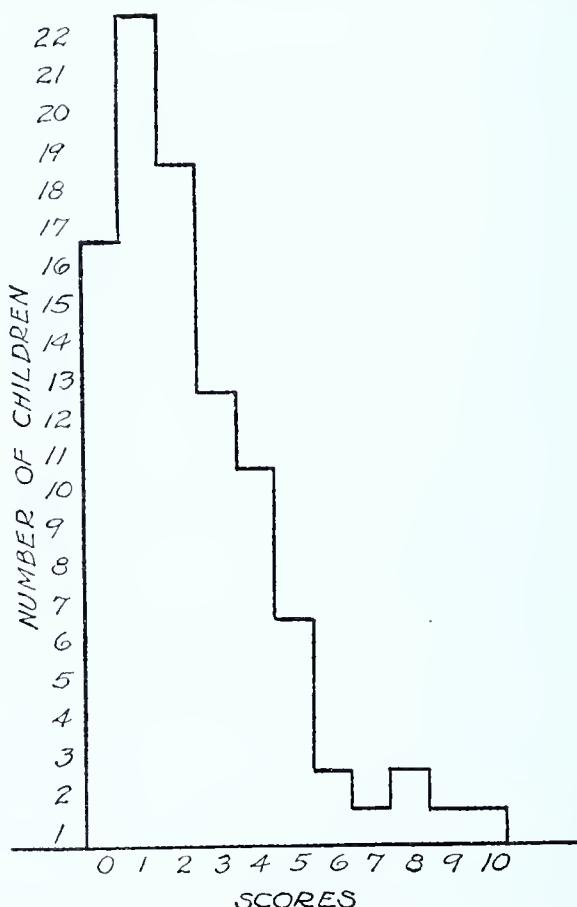


CHART XV. Distribution of Scores of 91 Three-Year-Olds in Use of Test Materials for Purposes Other Than Those Indicated

It is particularly interesting that the mean I.Q. of the group above the median in this table is higher than the mean I.Q. of the group below the median and this difference is statistically significant. This difference, however, to a great extent disappears when M.A. alone is considered. Chronological age, held constant in the comparison of I.Q.'s but not when M.A.

TABLE 16B

MEAN I.Q. AND M.A. OF GROUP ABOVE AND BELOW MEDIAN SCORE FOR "USE OF TEST MATERIALS OTHER THAN THOSE INDICATED," OMITTING FROM COMPUTATIONS 18 CASES FALLING AT MEDIAN

	ABOVE MEDIAN	BETWEEN MEDIAN	MEAN DIFFER- ENCE	P.E. OF DIFFER- ENCE
Mean I.Q. (Kuhlmann-Binet) .....	109	98	11	2.7
Mean I.Q. (Merrill-Palmer) .....	103.2	99.5	3.7	2.5
Mean M.A. (Kuhlmann-Binet) .....	39.1	36.9	2.2	1.0
Mean M.A. (Merrill-Palmer) .....	37.5	35.3	2.2	.9

alone was considered, is, despite its limited range of only six months, probably affecting this difference.

The mean I.Q. and M.A. of the 18 cases actually falling at the median were also computed. The mean I.Q. (Kuhlmann-Binet) is 112. The mean M.A. (Kuhlmann-Binet) is 40. The mean M.A. (Merrill-Palmer) is 37.7. In each case the mean scores are higher even than the mean scores of the group falling above the median in Table 16B. The number of cases is too small to warrant further statistical comparison, but the results undoubtedly indicate that any future studies that may be made on a larger number of cases, in respect to this behavior response as well as to the others, might profitably use comparisons more refined than the simple one of dividing the group above and below the median point. Moreover, as repeatedly indicated throughout this study, before this and other measures can be of any great use, studies must also be made of the appearance of these behavior items at various chronological age levels.

No statistical analysis, using the mean as the basis for dividing the group, has been made, since the mean score falls in the same class interval (class interval of 1) as the median.

## CHAPTER X

### SEX DIFFERENCES

It has been noted throughout this study, in our consideration of the various behavior items, that these items seem to exert little or no influence on the mental test results. They do, nevertheless, definitely discriminate among the children, regardless of their implications for the intelligence test score. Moreover, since it has previously been shown that these differences are being reliably measured, a brief discussion of the behavior responses themselves is not amiss.

It will be remembered that the group on which this study was based was equally divided on the basis of sex. In this

TABLE 17  
SEX DIFFERENCES IN TERMS OF MEAN SCORES OF VARIOUS MEASURES STUDIED

	BOYS		GIRLS		MEAN DIFFER- ENCE	P.E. OF DIFFER- ENCE
	Mean	Standard Devia- tion	Mean	Standard Devia- tion		
I.Q. ....	105.4	16.	110.3	17.0	+ 4.9*	2.9
M.A. (Kuhlmann-Binet) ....	37.3	6.0	39.4	6.8	+ 2.1	.9
M.A. (Merrill-Palmer) ....	35.5	5.0	37.7	5.6	+ 2.5	.7
Resistance ....	8.1	8.0	7.0	7.6	- 1.1*	†
Persistence ....	231.9	107.5	236.9	90.0	+ 5.0	(14.4)
Spontaneous conversation ...	162.5	102.5	121.0	75	- 41.5	(12.8)
Activity ....	11.5	11.7	10.0	9.6	- 1.5	
Initiative ....	17.4	9.6	15.2	9.3	- 2.2	(1.3)

\* + sign in difference column indicates higher average score for girls; — sign indicates higher average score for boys.

† The distributions of resistance and activity scores are definitely not "normal"; hence the P.E. of the difference formula cannot be used. Where parentheses are used, distributions are of doubtful character. The P.E. of the difference has been calculated, but its significance is probably doubtful.

chapter the various factors studied are examined for possible sex differences. Table 17 shows the mean scores, the standard deviations, the mean differences, and the P.E. of these differences for the boys, as compared with those for the girls.

As was to be expected, few differences exist, and these can hardly be considered significant. Any statistical results considered in this study of course have to be evaluated in the light of the homogeneity of the group under observation—a homogeneity which would tend to minimize any statistical differences. Possibly it would not be amiss to suggest that there seems to be a slight trend toward slightly higher intelligence scores at this age for girls as compared with boys—an inclination quite in accord with previous psychological studies of children;<sup>1</sup> and that, similarly, spontaneous behavior responses, considered in terms of initiative, activity, and conversation, would tend to indicate that there is slightly more overt expression on the part of the boys as contrasted with the girls.

<sup>1</sup> Goodenough, Florence, *The Kuhlmann-Binet Tests for Children of Preschool Age*, p. 40. Terman, L. M., *The Measurement of Intelligence*, p. 69.

## CHAPTER XI

### INTERRELATIONSHIPS EXISTING AMONG THE VARIOUS MEASURES

Thus far the behavior responses which we have been studying have been examined in terms of their own distribution, their range, means, and standard deviations, as well as in terms of the mean intelligence scores of two groups divided above and below the median score for each type of response.

Since we have indicated that, regardless of results obtained in relation to the test score itself, the measures are themselves

TABLE 18  
COEFFICIENTS OF CORRELATION EXISTING AMONG THE VARIOUS VARIABLES

	M.A. (KUHLMANN-BINET)	M.A. (MERRILL-PALMER)	RESISTANCE	PERSISTENCE	SPONTANEOUS CONVERSATION	SPONTANEOUS ACTIVITY (GROSS)	SPONTANEOUSLY REPEATING TEST OR DEMANDING OTHER TEST MATERIALS	SPONTANEOUS USE OF TEST MATERIALS FOR PURPOSES OTHER THAN INDICATED
M. A. (Kuhlmann-Binet) . . . . .		.75	-.32	.10	.28	.04	.08	.14
M.A. (Merrill-Palmer) . . . . .	.75		-.13	.01	.20	-.10	.11	.23
Resistance . . . . .	-.32	-.13		.09	-.19	.21	.13	.22
Persistence . . . . .	.10	.01	.09		.06	-.03	-.20	-.20
Spontaneous conversation . . . . .	.28	.20	-.19	.06		.43	.45	.34
Spontaneous activity (gross) . . . . .	.04	-.10	.21	.03	.43		.54	.09
Spontaneously repeating test or demanding other test materials . . . . .	.08	.11	.13	-.20	.45	.54		.24
Spontaneous use of test materials for purposes other than indicated . . . . .	.14	.23	.22	-.20	.34	.09	.24	

reliable indices of behavior responses, a study of the interrelationships existing among the various items may be profitable.

Correlation coefficients (Pearson  $r$ ) were obtained for each measure in relation to every other measure. Table 18 indicates the relationships thus found to exist.

We have previously discussed the relationship of the intelligence tests to the other responses, in terms of mean differences. It will be noted that in each instance only slight relationships exist, except in the case of resistance (which has been previously noted) and of spontaneous conversation.

The relationship existing between the two mental tests, however, warrants comment. The Pearson  $r$  is .75. In a previous research study, carried on by the writer, at the Child Development Institute, Teachers College, Columbia University (unpublished study), the Pearson  $r$  between mental ages on the two tests was found to be .80. This study was based not only on a larger number of children, but also on a superior group, on the whole, both in terms of I.Q. and of the occupational status of the fathers. Perhaps even more important was the wider age range (18 to 50.5 months) of the group. Not only was the correlation coefficient higher, but the mean difference in M.A. was found to be 4.2 months in favor of higher ratings on the Kuhlmann-Binet test, as compared to 2.0 months in the present study, where the mean M.A. (Kuhlmann-Binet) is 38.0; mean M.A. (Merrill-Palmer) is 36.0. Moreover, when broken down into comparable age groups, the mean difference in mental age in the previous study was found to be 5.4 months, at three years of age. It was further noted, however, that the mean difference between the two test ratings was differential with respect to both chronological age and brightness. That is, greater differences between the two tests existed when the brighter children (as measured by the Kuhlmann-Binet test) were used as a basis for a comparison of ratings on the two scales than when such comparisons were based on the duller children. In the fact that the first group was definitely superior to the more nearly average group studied

in the present research, we probably have an explanation of the decrease in difference between the two scales in the present study. However, this difference still exists, and it tends in the same direction.

Returning to a discussion of Table 18, we find, on the whole, no very significant relationships. This is not true, however, when we consider the relationship existing between spontaneous conversation and other spontaneous behavior, such as activity, initiative interest in test materials, and use of test materials for purposes other than indicated. The Pearson  $r$  between spontaneous conversation and spontaneous activity is  $.43 \pm .06$ ; between spontaneous conversation and initiative interest in tests,  $.47 \pm .06$ ; between spontaneous conversation and spontaneous use of test materials,  $.34 \pm .06$ . The  $r$  between use of test materials for purposes other than those indicated and initiative interest in tests is  $.24 \pm .07$ . Taking into consideration again the relative smallness and homogeneity of the group, these correlation coefficients may probably justifiably be considered quite significant. It is the writer's suggestion that, quite possibly, we are revealing here, in terms of these various behavior responses, a type of uninhibited, extraverted personality.

The correlation of  $.54 \pm .05$  existing between spontaneous activity and initiative interest in test materials is a comparatively high one. This cannot, however, be considered particularly significant if we remember that spontaneous activity was measured by the number of times a child left the testing table. Since a child often left the table to get new materials, it is readily seen that this activity would have been checked in both columns, and any resulting correlation between the two would be spurious.

Eventually such results as have been reached might profitably be examined by means of the multiple and partial correlation technique. Unfortunately, the scope of this study does not warrant further statistical analysis.

## CHAPTER XII

### INDIVIDUAL DIFFERENCES

The previous chapters have been devoted to a consideration of various responses revealed in a psychological test situation in terms of their possible influence on the psychological test situation itself, and also in relation to the differences which they may show exist between a group possessing a given trait to a greater extent as compared with a group possessing the same trait to a less extent. We have also examined the relationships existing between these various measures.

In some cases, actual differences appeared in the averages of the two groups studied, but these differences were never large, and in no case were they sufficiently significant to justify conclusions about any one individual. This discussion has been concerned with the average tendencies of two groups, their average sex differences, and the interrelationships of the various factors. It seems to the writer, however, that any one of these characteristics is important to note in terms of the individual children concerned, particularly since we have shown that children definitely differ in respect to the degree to which they express these different behavior responses and, moreover, that our measure of them is reliable.

An easy way to present a child's ranking on these given traits, and one which furnishes a basis for comparable results, is to examine the possession of a given characteristic in terms of the child's percentile ranks. If in addition we picture these percentile ranks graphically, we are furnished with a comparatively lucid picture of several aspects of a child's personality.

It probably would be presumptuous to consider the results of studying 91 cases—even though they definitely fall within a narrow age range—as the basis for valid norms as such,

However, percentile rank graphs of four children are presented, based on these results, as a means of illustrating the ultimate value of determining the degree to which a given child possesses a given characteristic.

## ROBERT GREEN

TRAIT	PERCENTILE RANK
Spontaneous conversation .....	84.2
Resistance .....	93.3
Persistence .....	6.6
Activity .....	88.4
Initiative interest in test materials .....	78.3
Intelligence .....	98.0

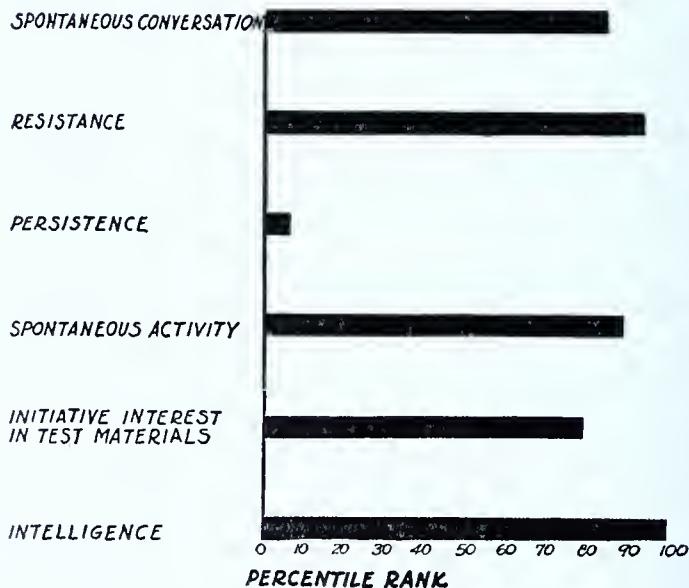


CHART XVI. Percentile Ranks in Various Traits of Robert Green

## EDITH MORLEY

TRAIT	PERCENTILE RANK
Spontaneous conversation	24.9
Resistance	50.1
Persistence	52.6
Activity	73.4
Initiative interest in test materials	28.1
Intelligence	51.9

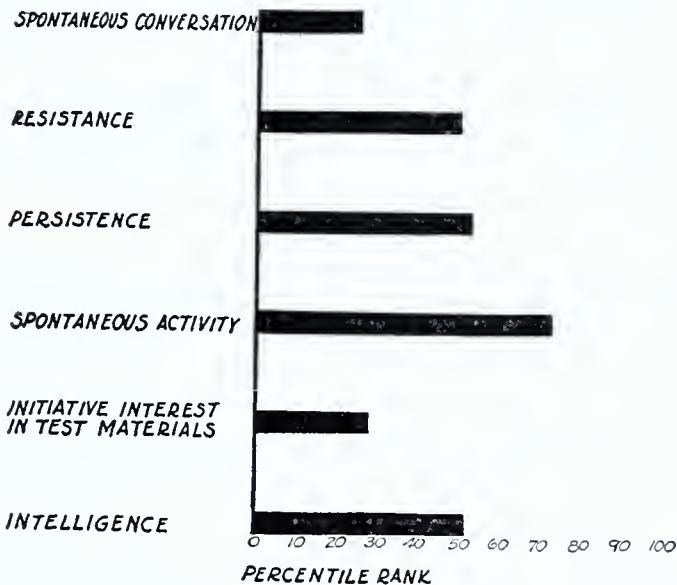


CHART XVII. Percentile Ranks in Various Traits of Edith Morley

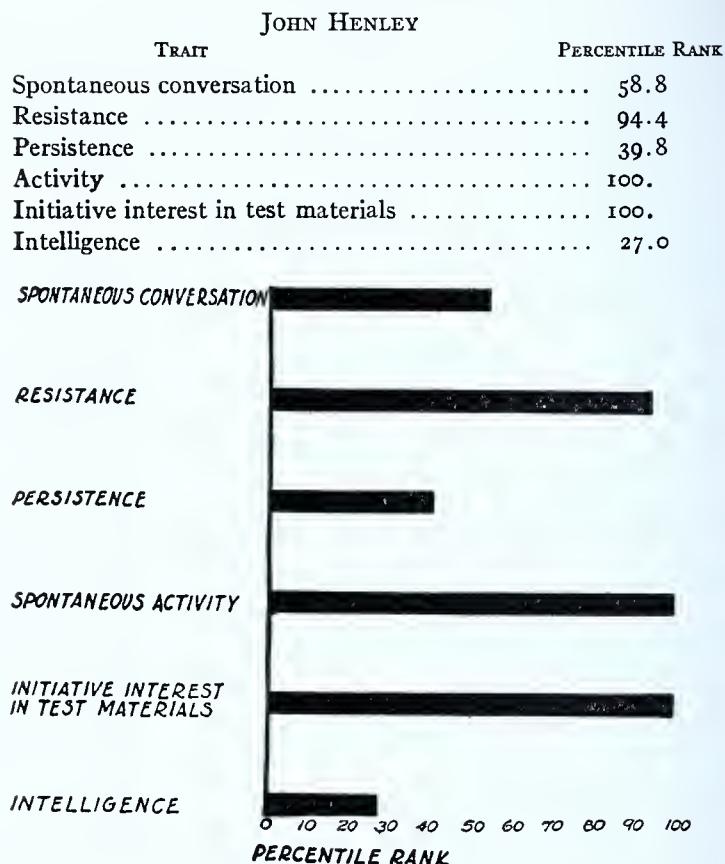


CHART XVIII. Percentile Ranks in Various Traits of John Henley

## LOUISE WATERS

TRAIT	PERCENTILE RANK
Spontaneous conversation	77.4
Resistance	50.6
Persistence	50.4
Activity	80.0
Initiative interest in test materials	33.7
Intelligence	80.9

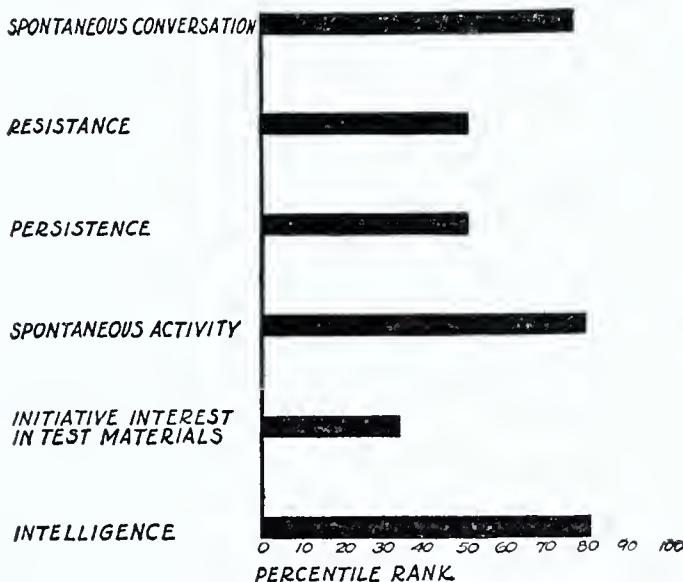


CHART XIX. Percentile Ranks in Various Traits of Louise Waters

To one familiar with the testing situation, the availability of such methods of describing a given child's reaction to a total situation, is significant.

It may be important for us to know that more intelligent three-year-olds tend, on the average, to be less resistant; although that fact, after all, tells us little or nothing about the individual child. He may have a high I.Q. He may also be very resistant. He may have little or no persistence as compared to other children of the same chronological age. Such comparisons are infinitely important to a complete understanding of the child as an individual; and if the results of this study contribute to establishing a method for a better evaluation of any given child's reactions, in other than purely intellectual terms, the research will have been justified.

## CHAPTER XIII

### SUMMARY

A method was developed whereby the psychological examiner during a simple mental test situation could record responses other than those called for by the test itself.

The behavior responses finally selected for study were: resistance, persistence, gross amount of spontaneous conversation, spontaneous activity, spontaneously repeating or demanding other test material, and the spontaneous use of test materials for purposes other than indicated at a given moment.

The reliability of recording these responses was found, in a preliminary study, to be uniformly very high. Rank order correlation coefficients between the examiner's records and the check observer's records ranged from .97 to 1.00.

The reliability of the data themselves was also found to be high. Using the odd-even technique on the results obtained in the main part of the study, correlation coefficients were found to range from .58 to .94. The "spontaneous use of test materials for purposes other than those indicated" was least reliable ( $r = .58$ ). Initiative interest in test materials ( $r = .84$ ) was next lowest.

The behavior responses studied, with the exception of resistance and spontaneous conversation, show but slight relationship to the mental test situation. The relationship between M.A. and resistance is not only negative, but significant ( $r = -.32 \pm .06$  between Kuhlmann-Binet M.A. and resistance). The relationship between Kuhlmann-Binet M.A. and spontaneous conversation is both positive and significant ( $r = .28 \pm .06$ ;  $r = .20 \pm .07$  between spontaneous conversation and Merrill-Palmer M.A.).

Sex differences were found to be negligible except for a pos-

sible tendency for the girls to have slightly higher intelligence test scores than the boys. The mean I.Q. for girls is 110.3; that for boys, 105.4. The boys tend slightly to exceed the girls in the amount of overt behavior responses, such as spontaneous conversation, activity, and initiative. The mean score on spontaneous conversation for boys is 162.5; for girls it is 121. The mean score on spontaneous activity for boys is 11.5; for girls it is 10.0. The mean score in initiative for boys is 17.4; for girls it is 15.2.

The interrelationships of the various factors considered in this study were for the most part negligible. However, most interesting were the comparatively high correlation coefficients (considering the smallness and homogeneity of the group studied) between spontaneous conversation and gross activity of the group ( $r = .43 \pm .06$ ); between spontaneous conversation and spontaneously repeating the test or seeking other test materials ( $r = .45 \pm .06$ ); and between spontaneous conversation and spontaneous use of test materials for purposes other than indicated in test situation ( $r = .34 \pm .06$ ). It is possible that these various measures of spontaneous behavior indicate what might be interpreted as expressions of an extravert, as contrasted with an intravert, type of personality.

## BIBLIOGRAPHY

1. GOODENOUGH, FLORENCE. *The Kuhlmann-Binet Tests for Children of Pre-School Age.* Monograph Series, No. 2. University of Minnesota Press, 1928.
2. GOODENOUGH, FLORENCE. "The Reliability and Validity of the Wallin Peg Board." *Psychological Clinic*, October 1927.
3. HOLZINGER, KARL J. *Statistical Methods for Students in Education.* Ginn and Co., 1928.
4. KUHLMANN, FRED. *Handbook of Mental Tests.* Warwick and York, 1922.
5. PIAGET, JEAN. *Language and Thought of the Child.* Harcourt, Brace and Co., 1926.
6. REYNOLDS, MARTHA MAY. *Negativism of Pre-School Children.* Bureau of Publications, Teachers College, Columbia University, 1928.
7. STUTSMAN, RACHAEL. *Performance Tests for Children of Pre-School Age.* Genetic Psychology Monograph, 1926, Vol. 1, No. 1. Clark University, 1926.
8. TERMAN, L. M. (Editor). *Genetic Studies of Genius*, Vol. 1. Stanford University Press, 1926.
9. TERMAN, L. M. *The Measurement of Intelligence.* Houghton Mifflin Company, 1916.
10. THOMAS, DOROTHY SWAINE, AND ASSOCIATES. *Some New Techniques for Studying Social Behavior.* Child Development Monograph, No. 1. Bureau of Publications, Teachers College, Columbia University, 1929.
11. ——. Chapter by Jenkins, Lulu-Marie. "A Study of the Reliability of Stenographers' and Students' Reports of Psychological Tests."
12. ——. Chapter by Nelson, Janet Fowler. "A Preliminary Report on Some Uses of the Psychological Test Situation for Studying Personality Differences."

## VITA

JANET FOWLER NELSON, born March 19, 1901 in Bridgeport, Conn. A.B. Vassar College, 1922. M.A. Columbia University, 1925.

Demonstrator in applied psychology, Bryn Mawr College 1923-24; Assistant Psychologist, Vassar Euthenics Conference, Summer 1926. Research Assistant, Child Development Institute, Teachers College, Columbia University, 1924-25-26; Research Associate, 1927-28. Psychologist and Statistician, Boys' Club Study, New York University, 1929-30-31.

